



# Backtesting Value-at-Risk: A GMM Duration-Based Test

Christophe Hurlin, Gilbert Colletaz, Sessi Tokpavi, Bertrand Candelon

## ► To cite this version:

Christophe Hurlin, Gilbert Colletaz, Sessi Tokpavi, Bertrand Candelon. Backtesting Value-at-Risk: A GMM Duration-Based Test. 2008. halshs-00329495

**HAL Id: halshs-00329495**

**<https://shs.hal.science/halshs-00329495>**

Preprint submitted on 10 Oct 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Backtesting Value-at-Risk: A GMM Duration-Based Test

Bertrand Candelon<sup>†</sup>, Gilbert Colletaz<sup>‡</sup>, Christophe Hurlin<sup>‡</sup>, Sessi Tokpavi<sup>‡</sup>

<sup>‡</sup> *University of Orléans, Laboratoire d'Economie d'Orléans (LEO), France.*

*Corresponding author: christophe.hurlin@univ-orleans.fr*

<sup>†</sup> *Maastricht University, Department of Economics. The Netherlands*

June 2008

---

## Abstract

This paper<sup>1</sup> proposes a new duration-based backtesting procedure for VaR forecasts. The GMM test framework proposed by Bontemps (2006) to test for the distributional assumption (*i.e.* the geometric distribution) is applied to the case of the VaR forecasts validity. Using simple  $J$ -statistic based on the moments defined by the orthonormal polynomials associated with the geometric distribution, this new approach tackles most of the drawbacks usually associated to duration based backtesting procedures. First, its implementation is extremely easy. Second, it allows for a separate test for unconditional coverage, independence and conditional coverage hypothesis (Christoffersen, 1998). Third, feasibility of the tests is improved. Fourth, Monte-Carlo simulations show that for realistic sample sizes, our GMM test outperforms traditional duration based test. An empirical application for Nasdaq returns confirms that using GMM test leads to major consequences for the *ex-post* evaluation of the risk by regulation authorities. Without any doubt, this paper provides a strong support for the empirical application of duration-based tests for VaR forecasts.

*Key words:* Value-at-Risk, backtesting, GMM, duration-based test.

*J.E.L Classification :* C22, C52, G28

---

---

<sup>1</sup> The authors thank Enrique Sentana for comments on the paper as well as the participants of the "Methods in International Finance Network" 2008 congress in Barcelona. The paper was partially performed during the visit of Christophe Hurlin at Maastricht University via the visiting professorship program of METEOR. Usual disclaimers apply.

## 1 Introduction

The recent Basel II agreements have left the possibility for financial institutions to develop and apply their own internal model of risk management. The Value-at-Risk (VaR thereafter), which measures the quantile of the projected distribution of gains and losses over a target horizon, constitutes the most popular measure of risk. Consequently, regulatory authorities need to set up adequate *ex-post* techniques validating or not the amount of risk taken by financial institutions. The standard assessment method of VaR consists in backtesting or reality check procedures. As defined by Jorion (2007), backtesting is a formal statistical framework that consists in verifying if actual trading losses are in line with projected losses. This involves a systemic comparison of the history of model-generated VaR forecasts with actual returns and generally relies on testing over VaR violations (also called the Hit).

A violation is said to occur when *ex-post* portfolio returns are lower than VaR forecasts<sup>2</sup>. Christoffersen (1998) argues that a VaR with a chosen coverage rate of  $\alpha\%$  is valid as soon as VaR violations satisfy both the hypothesis of unconditional coverage and independence. The hypothesis of unconditional coverage means that the expected frequency of observed violations is precisely equal to  $\alpha\%$ . If the unconditional probability of violation is significantly higher than  $\alpha\%$ , it means that VaR model understates the portfolio's actual level

---

<sup>2</sup> Or the opposite of *VaR* if the latter is defined as a loss in absolute value.

of risk. The opposite finding of too few VaR violations would alternatively signal an overly conservative VaR measure. The hypothesis of independence means that if the model of VaR calculation is valid then violations must be distributed independently. In other words, there must not have any cluster in the violation sequence. As noted by Campbell (2007), the unconditional coverage property places a restriction on how often VaR violations may occur, whereas the independence property restricts the ways in which these violations may occur. But both assumptions are essential to characterize VaR forecast validity: only hit sequences that satisfy each of these properties (and hence the conditional coverage hypothesis) can be presented as evidence of a useful VaR model.

Even if the literature about conditional coverage is quite recent, various tests on independence and unconditional coverage hypotheses have already been developed (see Campbell, 2007 for a survey). Most of them directly exploit the violation process<sup>3</sup>. However another streamline of the literature uses the statistical properties of the duration between two consecutive hits. The baseline idea is that if the one-period ahead VaR is correctly specified for a coverage rate  $\alpha$ , then the durations between two consecutive hits must have a geometric distribution with a success probability equal to  $\alpha\%$ . On these grounds Christoffersen and Pelletier (2004) proposed a test of independence. The gen-

---

<sup>3</sup> For instance, Christoffersen's test (1998) based on Markov chain, the hit regression test of Engle and Manganelli (2004) relies on a linear auto-regressive model, or the tests of Berkowitz and al. (2005) based on tests of martingale difference.

eral idea of their duration-based backtesting test consists in specifying a duration distribution that nests the geometric distribution and allows for duration dependence, so that the independence hypothesis can be tested by means of simple likelihood ratio (LR) tests. As noted by Haas (2007), this general duration-based approach of backtesting sounds very appealing. It is easy to apply and provides a clear-cut interpretation of parameters. Nevertheless, it must be noted that one has to specify a particular distribution under the alternative hypothesis. Moreover, LR test turns out to suffer from the relative scarcity of violations: even with one year of daily returns, the associated series of durations is likely to be short, in particular for a 1% coverage rate (the value recommended by supervision authorities). Consequently duration-based backtesting methods have relatively small power for realistic sample sizes (Haas, 2007) and it even often happens that standard LR duration-based statistics cannot be computed<sup>4</sup>. For these reasons, actual duration-based backtesting procedures are not very popular among practitioners. However we show in this paper that it is possible to significantly improve these procedures.

Relying on the GMM framework of Bontemps and Meddahi(2005,2006) we derive test statistics similar to  $J$ -statistics based on particular moments defined by the orthonormal polynomials associated with the geometric distribution.

Also our duration-based backtest considers discrete lifetime distributions: we

---

<sup>4</sup> The LR test requires at least one non-censored duration and an additional possibly censored duration (i.e. two violations) to be implemented. As experienced by Berkowitz et al. (2005) with one year of trading days ( $T = 250$ ) and  $\alpha = 0.01$  the test can be computed only in six cases out of ten.

expect in particular that this leads to an improvement in the power and size of our test upon ones based on continuous approximations as for example in Christoffersen and Pelletier (2004). To sum up, the present approach appears to have several advantages. First, it provides an unified framework in which we can investigate separately the unconditional coverage hypothesis, the independence assumption and the conditional coverage hypothesis. Second, the optimal weight matrix of our test is known and does not have to be estimated. Third the GMM statistics can be numerically computed for almost all realistic backtesting sample sizes. Fourth, it benefits from a result of Bontemps (2006) and appears to be robust to parameter uncertainty. Fifth, in contrast with the LR tests, it does not impose a particular distribution under the alternative. Finally, some Monte-Carlo simulations indicate that for realistic sample sizes, our GMM test have good power properties when compared to other duration-based backtests.

The paper is organized as follows. In section 2, we present the main VaR assessment tests and more particularly the duration-based backtesting procedures. Section 3 presents our GMM duration-based test. In section 4 we present the results of various Monte Carlo simulations in order to illustrate the finite sample properties of the proposed test. In section 5 we realize an empirical application using daily Nasdaq returns. Finally the last section concludes.

## 2 Duration-Based Backtesting of Value-at-Risk

Let us denote  $r_t$  the return of an asset or a portfolio of assets at time  $t$ . The *ex-ante* VaR for a  $\alpha\%$  coverage rate denoted  $VaR_{t|t-1}(\alpha)$ , anticipated conditionally to an information set  $\Omega_{t-1}$  available at time  $t-1$ , is defined by:

$$\Pr[r_t < VaR_{t|t-1}(\alpha)] = \alpha \text{ where } t = 1, \dots, T. \quad (1)$$

Let  $I_t(\alpha)$  be the hit variable associated to the *ex-post* observation of a  $\alpha\%$  VaR violation at time  $t$ .

$$I_t(\alpha) = \begin{cases} 1 & \text{if } r_t < VaR_{t|t-1}(\alpha) \\ 0 & \text{else} \end{cases} \quad (2)$$

As suggested by Christoffersen (1998), VaR forecasts are valid if and only if the violation sequence  $\{I_t\}_{t=1}^T$  satisfies the following two assumptions:

- The unconditional coverage (UC) hypothesis: the probability of a violation must be equal to the  $\alpha$  coverage rate:

$$\Pr[I_t(\alpha) = 1] = \mathbb{E}[I_t(\alpha)] = \alpha. \quad (3)$$

- The independence (IND) hypothesis: VaR violations observed at two different dates for the same coverage rate must be independently distributed.  
  
In other words, past VaR violations do not hold information about current and future violations.

The UC hypothesis is a straightforward one: if the frequency of violations

observed over  $T$  periods, is significantly lower (respectively higher) than the nominal coverage rate  $\alpha$ , then the VaR overestimates (respectively underestimates) the risk. Nevertheless, the UC hypothesis shades no light on the possible dependence of VaR violations and a model which does not satisfy the IND hypothesis could then lead to some clustering in violations and therefore does not offer a proper framework to appreciate the risk<sup>5</sup>.

When the assumptions of UC and IND hypotheses are simultaneously valid then VaR forecasts are said to have a correct conditional coverage (CC thereafter). Under the CC assumption, the VaR violation process becomes a martingale difference:

$$\mathbb{E}[I_t(\alpha) \mid \Omega_{t-1}] = \alpha. \quad (4)$$

The sequence  $\{I_t\}_{t=1}^T$  of VaR violations should then be a random sample from a Bernoulli distribution with a probability of violation equal to  $\alpha$ . It results that the process consisting in the numbers of periods between two violations has a geometric distribution with no memory. More precisely, it signifies that when a violation occurs then the conditional probability distribution of the number of periods which are passed before a new hit occurs does not depend on how many violations were already observed.

---

<sup>5</sup> An efficient measure of risk should indeed adjust automatically and immediately to any new information so that the ex-ante probability of a violation for  $t+1$  must be equal to the accepted nominal coverage rate whatever the value of the hit function at time  $t$ .



Formally, let us denote  $D_i$  the duration between two consecutive violations as:

$$D_i = t_i - t_{i-1}, \quad (5)$$

where  $t_i$  denotes the date of the  $i^{th}$  violation. Under CC hypothesis, the duration  $D_i$  follows a geometric distribution with a probability equal to  $\alpha$  and a probability mass function given by:

$$f_D(d; \alpha) = (1 - \alpha)^{d-1} \alpha \quad d \in \mathbb{N}^*. \quad (6)$$

Exploiting this relation, it is straightforward to develop a likelihood ratio test for the IND and/or the CC hypotheses. The general idea of these duration-based backtesting tests consists in specifying a duration distribution nesting the geometric distribution  $f_D(d; \alpha)$  and allowing for duration dependence.<sup>6</sup>

Hence Christoffersen and Pelletier (2004) proposed the first duration-based test for which they used the exponential distribution, which is the continuous analogue of the geometric distribution, and has a probability density function defined as:

$$f_D(d; p) = p \exp(-pd). \quad (7)$$

With (7) we have  $\mathbb{E}(d) = 1/p$  and, as the CC hypothesis implies a mean duration equals to  $1/\alpha$ , it also implies the condition  $p = \alpha$ . With regard to the IND hypothesis, they postulate a Weibull distribution so that under the

---

<sup>6</sup> The memory-free property implies a flat hazard function. On the contrary, violation clustering corresponds to decreasing hazard function, implying that the probability of no violation spell ending decreases as the spell increases in length.

alternative the density of durations between successive hits is given by:

$$f_D(d; p, b) = p^b b d^{b-1} \exp \left[ - (p d)^b \right]. \quad (8)$$

As the Exponential is a Weibull with a flat hazard function, *i.e*  $b = 1$ , the test for IND (Christoffersen and Pelletier, 2004) is then simply:

$$H_{0,IND} : b = 1. \quad (9)$$

In a recent work, Berkowitz et al.(2005) extended this approach to consider the CC hypothesis, that is:

$$H_{0,CC} : b = 1, \quad p = \alpha. \quad (10)$$

However even if duration-based backtesting tests relying on continuous function are attractive because of their elegance and simplicity, they are not entirely satisfying. In particular, Haas (2005) motivates the use of discrete lifetime distributions instead of continuous ones, arguing that the parameters of the distribution have a clear-cut interpretation in terms of risk management. He also conducts Monte-Carlo experiments showing that the backtesting tests based on discrete distribution exhibit a higher power than the continuous competitor tests.

Moreover, and independently of this aspect, other limitations may explained the lack of popularity of duration-based backtesting tests among practitioners. First, they exhibit low power for realistic backtesting sample sizes. For instance, in some GARCH based experiments Haas (2005) founds that for

a backtesting sample size of 250, the LR independence tests have an effective power that ranges from 4.6% (continuous Weibull test) to 7.3% (discrete Weibull test) for a nominal coverage of 1%VaR. Similarly, for a coverage of 5%VaR, the power only reaches 14.7% for the continuous Weibull test and 32.3% for the discrete Weibull test. In other words, when VaR forecasts are not valid, LR tests do not reject the VaR validity at best in 7 cases out of 10. Similar lack of power is also apparent in Hurlin and Tokpavi (2007).

Second, duration based tests turn out to have a rather limited feasibility. Indeed, for realistic backtesting sample sizes ( $T$  around 250) and a coverage rate of 1%, it often happens that the LR duration-based statistics cannot be computed. This is because the implementation of the test requires at least one non-censored duration and an additional, possibly censored, duration (*i.e.* two violations). As observed by Berkowitz et al. (2005), this call for rather huge samples: when he considers one year of trading days ( $T = 250$ ) then, under the null of conditional coverage with  $\alpha = 0.01$ , the Weibull LR test can be computed only in 6 cases out of 10. In other words, in 4 cases out of 10, the duration based statistics can not be used to test the VaR forecasts validity, whereas other backtesting approaches not based on durations<sup>7</sup> can be used.

Third, duration-based tests do not allow formal separate tests for UC, IND and CC within a unified framework<sup>8</sup>. It seems however usefull to propose a

---

<sup>7</sup> as in Kupiec (1995), Christoffersen (1998), Engle and Manganelli (2004).

<sup>8</sup> A characteristic which they share with other approaches (Christoffersen, 1998 or Engle and Manganelli, 2004).

backtesting strategy that could test (i) the unconditional coverage, the (ii) conditional coverage assumption and eventually (iii) the independence assumption.

### 3 A GMM Duration-Based Test

This paper proposes a new duration-based backtesting tests able to tackle these three issues. Extending the framework proposed by Bontemps and Meddahi, (2005, 2006) and Bontemps (2006), it consists in using a GMM framework in order to test if durations of VaR violations are geometrically distributed. This approach presents several advantages. First, it is extremely easy to implement, as it consists in GMM moment condition test. Second, it requires only few constraints improving the feasibility of the backtesting tests. Third, it allows for an optimal treatment of the problem associated with parameter uncertainty. Finally the choice of moment conditions enables us to elaborate separate tests for the UC, IND and CC assumptions, which was not possible with the existing duration based tests. Finally, Monte-Carlo simulations will show that this new test has relatively high power properties.

#### 3.1 Orthonormal Polynomials and Moment Conditions

In the continuous case, it is well known that the Pearson family of distributions (Normal, Student, Gamma, Beta, Uniform..) can be associated to some particular orthonormal polynomials whose expectation is equal to zero. These

polynomials can be used as special moments to test for a distributional assumption. For instance, the Hermite polynomials associated to the normal distribution are employed to test for normality (Bontemps and Meddahi, 2005). In the discrete case, orthonormal polynomials can be defined for distributions belonging to the Ord's family (Poisson, Binomial, Pascal, hypergeometric). The orthonormal polynomials associated to the geometric distribution (6) are defined<sup>9</sup> as follows:

**Definition 1** *The orthonormal polynomials associated to a geometric distribution with a success probability  $\beta$  are defined by the following recursive relationship,  $\forall d \in \mathbb{N}^*$ :*

$$M_{j+1}(d; \beta) = \frac{(1 - \beta)(2j + 1) + \beta(j - d + 1)}{(j + 1)\sqrt{1 - \beta}} M_j(d; \beta) - \left(\frac{j}{j + 1}\right) M_{j-1}(d; \beta), \quad (11)$$

for any order  $j \in \mathbb{N}$ , with  $M_{-1}(d; \beta) = 0$  and  $M_0(d; \beta) = 1$ . If the true distribution of  $d$  is a geometric distribution with a success probability  $\beta$  then, it follows that:

$$\mathbb{E}[M_j(d; \beta)] = 0 \quad \forall j \in \mathbb{N}^*, \forall d \in \mathbb{N}^*. \quad (12)$$

The duration GMM backtesting procedure exploits these moment conditions. More precisely, let us define  $\{d_1, \dots, d_N\}$  a sequence of  $N$  durations between two consecutive VaR violations observed over  $T$  periods. Under the conditional coverage assumption, the durations  $d_i$ ,  $i = 1, \dots, N$ , are *i.i.d.* and has a

---

<sup>9</sup> These polynomials can be viewed as a particular case of the Meixner orthonormal polynomials associated to a Pascal (negative Binomial) distribution.

geometric distribution with a success probability equals to the coverage rate  $\alpha$ . Hence, the hypothesis of a correct conditional coverage shortfall probability  $\alpha$  can be expressed as follows:<sup>10</sup>

$$H_{0,CC} : \mathbb{E}[M_j(d_i; \alpha)] = 0 \quad j = \{1, \dots, p\}, \quad (14)$$

where  $p$  denotes the number of moment conditions, with  $p > 1$ .

This framework also allows to test separately for the UC hypothesis. Under UC, the mean of durations between two violations is equal to  $1/\alpha$ , and this null hypothesis for UC can then be expressed as<sup>11</sup>:

$$H_{0,UC} : \mathbb{E}[M_1(d_i; \alpha)] = 0. \quad (15)$$

Thus, any discrete distribution satisfying the property  $\mathbb{E}[M_1(d; \alpha)] = 0$ , respects the UC hypothesis, whatever its behavior in term of dependence. For such a reason, this test can be interpreted as a simple unconditional coverage test.

Finally, a separate test for the IND hypothesis can also be derived. It consists in testing the hypothesis of a geometric distribution (implying the absence of

---

<sup>10</sup> It is possible to test the conditional coverage assumption by considering at least two moment conditions even if they are not consecutive as soon as the first condition  $\mathbb{E}[M_1(d_i)] = 0$  is included in the set of moments. For instance, it is possible to test the CC with:

$$H_{0,CC} : \mathbb{E}[M_j(d_i)] = 0 \quad j = \{1, 3, 7\} \quad (13)$$

For simplicity, we exclusively consider in the rest of the paper the cases where moment conditions are consecutive polynomials.

<sup>11</sup> Indeed, since  $M_1(d; \alpha) = (1 - \alpha d) / \sqrt{1 - \alpha}$ , it is straightforward to verify that the condition  $\mathbb{E}[M_1(d; \alpha)] = 0$  is equivalent to the UC condition  $\mathbb{E}(d) = 1/\alpha$ .

dependence) with a success probability equal  $\beta$ , where parameter  $\beta$  can be either fixed *a priori*, either estimated and is not necessarily equal to the coverage rate  $\alpha$ . This independence assumption can be expressed as the following moment conditions:

$$H_{0,IND} : \mathbb{E}[M_j(d_i; \beta)] = 0 \quad j = 1, \dots, p, \quad (16)$$

with  $p > 1$ . In this case, the average duration  $\mathbb{E}(d)$  is equal to  $1/\beta$  as soon as the first polynomial  $M_1(d; \beta)$  is included in the set of moments conditions. So, under  $H_{0,IND}$ , the durations between two consecutive violations have a geometric distribution and the *UC* is not valid if  $\beta \neq \alpha$ .

### 3.2 Empirical Test Procedure

It turns out that VaR forecast tests can be expressed as simple moment conditions, which can be tested within the well-known GMM framework. The philosophy of the test is to choose the appropriate moments and to test if their empirical expectations are close to 0 or not. As observed by Bontemps (2006), the orthonormal polynomials  $M_j(d; \beta)$  present two great advantages. First, the corresponding moments are robust to parameter uncertainty and second, the asymptotic matrix of variance covariance is known. Considering the last point it appears that in an *i.i.d.* context the moments are asymptotically independent with unit variance. As a consequence, the optimal weight matrix of the GMM criteria is simply an identity matrix and the implementation of the backtesting test becomes very easy.

**Proposition 2** *For a given model  $\mathcal{M}$  and a fixed coverage rate  $\alpha$ , let us consider a sequence of  $N$  durations, denoted  $\{d_i\}_{i=1}^N$ , observed between two successive violations associated to the  $\alpha\%$  VaR forecasts. The null hypothesis of CC can be expressed as:*

$$H_{0,CC} : \mathbb{E} [M (d_i; \alpha)] = 0, \quad (17)$$

where  $M (d_i; \alpha)$  denotes a  $(p, 1)$  vector whose components are the orthonormal polynomials  $M_j (d_i; \alpha)$ , for  $j = 1, \dots, p$ . Under some regularity conditions, we know since Hansen (1982) that

$$\left( \frac{1}{\sqrt{N}} \sum_{i=1}^N M_j (d_i; \alpha) \right)^2 \xrightarrow[N \rightarrow \infty]{d} \chi^2 (1) \quad , \forall j = 1, \dots, p, \quad (18)$$

So that, in an i.i.d. context these moments are asymptotically independent with unit variance and the conditional coverage (CC) statistic test is

$$J_{CC} (p) = \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N M (d_i; \alpha) \right)^\top \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N M (d_i; \alpha) \right) \xrightarrow[N \rightarrow \infty]{d} \chi^2 (p), \quad (19)$$

with  $p$  is the number of orthonormal polynomials used as moment conditions.

The  $J_{CC} (p)$  test statistic is easy to compute and follows a standard asymptotic distribution.

Test statistic for UC, denoted  $J_{UC}$ , is obtained as a special case of the proposition (2), when one considers only the first orthonormal polynomial, i.e. when  $M (d_i; \alpha) = M_1 (d_i; \alpha)$ .  $J_{UC}$  is then equivalent to  $J_{CC} (1)$  and can be expressed as follows:



$$J_{UC}(p) = \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N M_1(d_i; \alpha) \right)^2 \xrightarrow[N \rightarrow \infty]{d} \chi^2(1). \quad (20)$$

Finally, the statistic for  $IND$ , denoted  $J_{IND}$ , is defined for a success probability  $\beta$  and then can be expressed as follows

$$J_{IND}(p) = \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N M(d_i; \beta) \right)^\top \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N M(d_i; \beta) \right) \xrightarrow[N \rightarrow \infty]{d} \chi^2(p). \quad (21)$$

where  $M(d_i; \beta)$  denotes a  $(p, 1)$  vector whose components are the orthonormal polynomials  $M_j(d_i; \beta)$ , for  $j = 1, \dots, p$ , evaluated for a success probability equal to  $\beta$ .

### 3.2.1 Remark 1: Parameter Uncertainty

The moment-based tests raise a potential problem of parameter uncertainty (Bontemps, 2006; Bontemps and Meddahi 2006), when some parameters of the distribution under the null are unknown and must be estimated. Concerning our framework, it is obvious that the tests for CC and UC do not face such a problem. In these cases, the only parameter, *i.e.*  $\alpha$ , is known, since it represents the coverage rate defined *ex-ante* by market regulators or risk managers. So,  $J_{CC}$  and  $J_{UC}$  are thus free of parameter uncertainty. On the contrary, the test for IND hypothesis might be subject to uncertainty problem as the true VaR violations rate  $\beta$  is unknown.<sup>12</sup> Consequently, the independence test statistic,

---

<sup>12</sup> The true VaR violations rate may be different from the coverage rate  $\alpha$  fixed by the risk manager in the model.

$J_{IND}(p)$ , must be based on moments that depend on estimated parameters, *i.e.* instead of having  $\mathbb{E}[M_j(d_i; \beta)]$  where  $\beta$  is known we have to deal with  $\mathbb{E}[M_j(d_i; \hat{\beta})]$  where  $\hat{\beta}$  denotes a square- $N$ -root-consistent estimator<sup>13</sup> of  $\beta$ . It is well known that replacing the true value of  $\beta$  by its estimates  $\hat{\beta}$  may change the asymptotic distribution of the GMM statistic. However, Bontemps (2006) shows that the asymptotic distribution remains unchanged if the moments can be expressed as a projection onto the orthogonal of the score. Appendix A shows that the moment conditions defined by the Meixner orthonormal polynomials satisfy this property. So, the moments used to define the  $J_{IND}(p)$  statistic are robust against the problem of the parameter uncertainty and the asymptotic distribution of the GMM statistic  $J_{IND}(p)$ , based on  $M_j(d_i; \hat{\beta})$ , is similar to the one based on  $M_j(d_i; \beta)$ . So, we have:

$$J_{IND}(p) = \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N M(d_i; \hat{\beta}) \right)^{\top} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N M(d_i; \hat{\beta}) \right) \xrightarrow[N \rightarrow \infty]{d} \chi^2(p-1), \quad (22)$$

where  $M(d_i; \hat{\beta})$  denotes the  $(1, p)$  vector defined as  $(M_1(d_i; \hat{\beta}) \dots M_p(d_i; \hat{\beta}))$ . Note that in this case, the first polynomial  $M_1(d_i; \hat{\beta})$  is strictly proportional to the score used to defined the maximum likelihood estimator  $\hat{\beta}$  and thus  $M_1(d_i; \hat{\beta}) = 0$ . So, the degree of freedom of the  $J$ -statistic has to be adjusted accordingly.

---

<sup>13</sup> In our applications, we consider the *ML* estimator of  $\beta$ .

### *3.2.2 Remark 2: Small Sample Property*

One of the main issues in the literature on VaR assessment is the relative scarcity of violations. As recalled previously, even with one year of daily returns the number of observed durations between two hits may often be dramatically small, in particular for a 1% coverage rate. This may induce small sample bias that can be corrected with bootstrap experiments but some precautions must be taken. While the statistic used to test the unconditional coverage hypothesis is pivotal, this is not the case for the statistic associated with the test of the independence assumption as it depends on  $\hat{\beta}$ . For this reason, the size of this test has to be controlled using for example the Monte Carlo testing approach of Dufour (2006), as done for example in Christoffersen and Pelletier (2004).

### *3.3 Simulation Framework for Empirical Size Analysis and Numerical Aspects*

To illustrate the size performance of our duration-based test in finite sample, Monte-carlo experiments were performed. To generate hits sequence of violations we take independent draws from a Bernoulli distribution, considering successively  $\alpha = 1\%$  and  $\alpha = 5\%$  for the VaR nominal coverage. Several sample sizes  $T$ , ranging from 250 (which roughly corresponds to one year of trading days) to 1,500 were also used. Reported empirical sizes correspond to the rejection rates calculated over 10,000 simulations for a nominal size fixed

at 10%. If the asymptotic distribution of our test is adequate, the rejection frequency should be around the nominal size. The test for CC proposed by Berkowitz et al. (2005) (thereafter labeled  $LR_{CC}$ ) constitutes the benchmark against which we appreciate the properties of our GMM backtesting test.

*Insert Table 1*

The rejection frequencies of the Monte-Carlo experiments are presented in Table 1. It turns out that whatever  $\alpha$  our test is undersized in finite sample, but converges to the nominal size when  $T$  increases. However, recall that under the null in a sample with  $T = 250$  and a coverage rate equal to 1%, the expected number of durations between two consecutive hits ranges between two and three. This scarcity of violations explains why the empirical size of our asymptotic test is different from the nominal size in small samples. For this reason, in the next sections, we will use the Monte Carlo testing approach of Dufour (2006) that allows to control for the size even in small samples. It also appears that under-rejection worsens as the number of moment conditions used increases. On the contrary, we verify that the LR test proposed by Berkowitz et al. (2005) is oversized.

However, it is important to note that these rejection frequencies are only calculated for the simulations providing a  $J_{CC}$  as well as the  $LR_{CC}$  test statistics. Indeed, for realistic backtesting sample size (for instance  $T = 250$ ) and a coverage rate of 1%, many simulations do not deliver a statistics. As

previously mentioned, the  $LR_{CC}$  test requires to be implemented at least one non-censored duration and an additional possibly censored duration (*i.e.* two violations). By comparison, the implementation of our GMM test only requires at least one violation (*i.e.* one or two censored durations). Table 2 reports the feasibility ratios, *i.e.* the fraction of simulated samples where the  $LR_{CC}$  and the  $J_{CC}$  tests are feasible.

*Insert Table 2*

As observed by Berkowitz et al. (2005), these results highlight huge differences in the cases of 1% VaR and samples of 250 – 500 observations. For instance, when we consider one year of trading days ( $T = 250$ ), under the null of conditional coverage, the Weibull  $LR_{CC}$  test can be computed only in 6 samples out of 10. Such a simulation exercise illustrates one the advantages of our GMM duration-based test. For higher sample size (*i.e.* two years of trading days,  $T = 500$ ), the feasibility ratio is similar for both tests and lies around 1.

### *3.4 Simulation Framework for Power Analysis*

We now investigate the power of the test for different alternative hypothesis. Following Christoffersen and Pelletier (2004), Berkowitz et al.(2005) or Haas (2005), the DGP under the alternative hypothesis assume that returns,  $r_t$ , are issued from a GARCH(1, 1) –  $t(d)$  model with an asymmetric leverage effect.

More precisely, it corresponds to the following model:

$$r_t = \sigma_t z_t \sqrt{\frac{v-2}{v}}, \quad (23)$$

where  $\{z_t\}$  is an *i.i.d.* sequence form a Student's t-distribution with  $v$  degrees of freedom and where conditional variance is given by:

$$\sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 \left( \sqrt{\frac{v-2}{v}} z_{t-1} - \theta \right)^2 + \beta \sigma_{t-1}^2. \quad (24)$$

Parametrization of the coefficients is also similar to the one proposed by Christoffersen and Pelletier (2004) and used by Haas (2005), *i.e.*  $\alpha = 0.1$ ,  $\theta = 0.5$ ,  $\beta = 0.85$ ,  $\omega = 3.9683e^{-6}$  and  $d = 8$ . The value of  $\omega$  is set to target an annual standard deviation of 0.20 and the global parametrization implies a daily volatility persistence of 0.975.

Using the simulated Profit and Loss (P&L thereafter) distribution issued from this DGP, it is then necessary to select a method to forecast the VaR. This choice is of major importance for the power of the test. Indeed, it is necessary to choose a VaR calculation method which are not adapted to the P&L distribution and therefore violate efficiency, *i.e.* the nominal coverage and/or independence hypothesis. Of course, we expect that the larger the deviation from the nominal coverage and/or independence hypothesis will be, the higher the power of the tests will become. For comparison purpose, we consider the VaR calculation method used by Christoffersen and Pelletier (2004), Berkowitz et al.(2005) or Haas (2005), *i.e.* the Historical Simulation (HS). As in Christoffersen and Pelletier (2004), the rolling windows  $Te$  is taken to be either 250

or 500. Formally, HS-VaR is defined by the following relation:

$$VaR_{t|t-1}(\alpha) = percentile \left( \{r_j\}_{j=t-T_e}^{t-1}, 100\alpha \right). \quad (25)$$

HS easily generates VaR violations. In Figure 1, observed simulated returns  $r_t$  for a given simulation and VaR-HS are plotted. It appears that violation clusters are evident, whether for 1% VaR or for 5% VaR .

*Insert Figure 1*

For each simulation, the zero-one hit sequence  $I_t$  is calculated by comparing the *ex post* returns  $r_t$  to the *ex ante* forecast  $VaR_{t|t-1}(\alpha)$ , and the sequence of durations  $Di$  (or  $Y_i$ ) between violations are calculated from the hit sequence. From this duration sequence, the test-statistics  $J_{CC}(p)$  for  $p = 1, \dots, 5$  as well as the Berkowitz et al. (2005) test ( $LR_{CC}$ ) are implemented. The empirical power of the tests is then deduced from rejection frequencies based on 10,000 replications. However, as previously mentioned, the use of asymptotic critical values (based on a  $\chi^2$  distribution) induces important size distortions even for relatively large sample. So given the scarcity of violations (particularly for a 1% coverage rate), it is particularly important to control for the size of the backtesting tests. As usual in this literature, the Monte Carlo technique proposed by Dufour (2006) is implemented (See Appendix B).

*Insert Tables 3 and 4*

Tables 3 and 4 reports the rejection frequencies (power) of the test for respec-

tively 1% and 5% VaR.<sup>14</sup> We report the power of our test for various values of the number of moment conditions,  $p$ . We can observe that for small  $p$  values, the power is increasing with  $p$ . This result illustrates the fact that the Bon-temps's framework is not robust to any specification under the alternative if one uses only a few number of polynomials. Each test based on a specific polynomial is robust against the alternatives for which the corresponding moment has some expectation different from zero. Therefore the tests will be robust only if we consider a sufficient number of polynomials. In our simulations, it turns out that the power is optimal when considering three moment conditions in the case of the 1% VaR whereas five Meixner polynomials are required for a 5% VaR. To illustrate this point, the power is plotted for different number of moment conditions in Figure 2.

*Insert Figure 2*

In all cases the power of the GMM based backtesting test  $J_{CC}$  is greater than the one of the Berkowitz et al (2005) test whatever the sample size considered. In particular, the gain of our test is specially noticeable for the more interesting cases from a practical point of view, that is small sample size and  $\alpha = 1\%$ . For  $T = 250$ , the power of our test is two times the power of standard LR test. Such a property constitutes a key point to promote the empirical popularity of duration based backtesting tests. The comparison

---

<sup>14</sup> We only report results for the UC and CC. Outcomes of the simulations for IND tests are available upon request from the authors.



of the test for UC is impossible as traditional tests do not provide such an information.<sup>15</sup> Nevertheless, its power is always relatively high and in all case larger than 25%.

These simulations experiments confirm that GMM based duration test improves the feasibility and the power of traditional duration based tests. Besides it provides a separate test for CC, UC and IND hypotheses. Our initial objectives are thus fulfilled.

## 4 Empirical Application

To illustrate these new tests, an empirical application is performed, considering three sequences of 5%VaR forecasts on the daily returns of the Nasdaq index. These sequences correspond to three different VaR forecasting methods traditionally used in the literature: a pure parametric method (GARCH model under Student distribution), a non parametric method (Historical Simulation) and a semi parametric method based on a quantile regression (CAViaR, Engle and Manganelli, 2004). Each sequence contains 250 successive one-period-ahead forecasts for the period June 22, 2005 to June 20, 2006. The parameters of the GARCH and CAViaR models are estimated according to a rolling windows method with a length<sup>16</sup> fixed to 250 observations.

---

<sup>15</sup> As already noticed, traditional duration based tests do not provide a separate test for UC.

<sup>16</sup> The total sample runs from June 20, 2004 to June 20, 2006 (500 observations). The length of the rolling estimation window of the HS is also fixed to 250 observations.

The observed returns and 5%VaR forecasts obtained using the three alternative computation methods are reported on Figure C.3. As usually, it can be checked that the HS VaR forecasts are relatively flat. This result is fully intuitive since HS-VaR is calculated as the unconditional quantile of past returns: the time variability is then only captured via the rolling historical sample. On the contrary, the VaRs based on conditional variance or conditional quantile are more flexible. Consequently, as it is generally observed in the literature, the HS-VaR is more likely to generate violations clusters than other methods. Figure C.4 displays the indicator variable  $I_t(\alpha)$  associated to the ex-post 5%VaR violation computed for the three methods. As usual, HS-VaR exhibits clustering in violations: six out of the nine VaR-HS violations occur at the end of the period. The clusters are less obvious when considering the others methods. We also observe on both Figures that the VaR computed from CAViaR is clearly too low compared to the historical returns, i.e. the risk is underestimated: it leads to only seven violations over a sample of 250 observations implying a hit frequency rate around 2.8%. By comparison, there are nine hits for GARCH and HS methods, implying a frequency hit rate equal to 3.6%. So, in this configuration, if rejection of UC occurs, we expect that it occurs for the VaR-CAViaR. If rejection of IND occurs, we expect that it happens for HS-VaR.

The results obtained using the GMM duration-based tests are reported in Table 5. For each VaR method, we report the UC, CC and IND statistics. For

the two last tests, the number of moments  $p$  is fixed to 2, 4 and 6. For seek of comparison  $LR_{CC}$  statistics (Christoffersen and Pelletier, 2004; Berkowitz et al. 2005) are also reported. For all tests, the  $p$ -values correspond to the size-corrected ones (Dufour, 2006). Several comments can be done on these results.

First, our unconditional coverage test statistic  $J_{UC}$  leads to an unambiguous rejection of the validity of the CAViaR based VaR. As expected, this results is due to the too low violation rate associated to this method. Of course, the value of  $J_{UC}$  is identical for HS and GARCH, since these two methods lead to the same number of hits even if these violations do not occur at the same periods. Second, we observe that our GMM independence test ( $J_{IND}$ ) is able to reject (except in the case  $p = 2$ ) the null for HS-VaR. On the contrary,  $LR_{IND}$  test does not reject the null of independence for any of the three VaRs. Third, at a 10% significance level, the GMM conditional coverage test ( $J_{CC}$ ) rejects the validity of CAViaR and HS VaR<sup>17</sup> forecasts, contrary to standard LR tests. Finally, the GARCH-t( $d$ ) turns out to be best way to forecast risk: the UC, IND and CC are not rejected.

---

<sup>17</sup> At a 5% significance level, the null is rejected for  $p = 4$  and  $p = 6$ . When  $p$  is equal to 2 the  $p$ -value is equal to 0.15.

## 5 Conclusion

This paper develops a new duration-based backtesting procedure for VaR forecasts. The underlying idea is that if the one-period ahead VaR is correctly specified, then, every period, the duration until the next violation should be distributed according to a geometric distribution with a success probability equal to the VaR coverage rate. So, we adapt the GMM framework proposed by Bontemps (2006) in order to test for this distributional assumption that corresponds to the null of VaR forecast validity. The test statistics boils down to a simple  $J$ -statistic based on particular moments defined by the orthonormal polynomials associated to the geometric distribution. This new approach tackles most of the drawbacks usually associated to duration based model. First, its implementation is extremely easy. Second, it allows for a separate the unconditional coverage, the independence and the conditional coverage hypothesis (Christoffersen, 1998). Second, feasibility of the tests is improved. Third, Monte-Carlo simulations show that for realistic sample sizes, GMM test outperforms traditional duration based test. Our empirical application for Nasdaq returns confirms that using GMM test leads to major for the *ex-post* evaluation of the risk by regulation authorities. Our hope is that this paper will constitute an incitation for regulation authorities in order to use of duration-based tests to assess the risk taken by financial institutions. There is no doubt that a more adequate evaluation of the risk would decrease the probability of banking crises and systemic banking fragility.

Nevertheless, we have to admit that this test does not constitute the cure to all disease and several limits associated to duration based test are still to be tackled. In particular, it turns out that the estimation of the risk is neglected leading to severe bias (Escanciano and Olmo, 2007) . This will constitute a direct extension for the GMM approach and already lies among our future research plans.

## A Appendix: Proof of parameter uncertainty robustness of $J_{IND}$

Under the independence assumption, the sequence of durations  $\{d_i\}_{i=1}^N$  is a sequence of *i.i.d.* geometric random variables with a success probability  $\beta$ , where  $\beta$  is *a priori* unknown. The probability distribution function of  $d_i$  is:

$$f_D(d; \beta) = (1 - \beta)^{d-1} \beta \quad d \in \mathbb{N}^*. \quad (\text{A.1})$$

The score is then defined as:

$$\frac{\partial \ln f_D(d; \beta)}{\partial \beta} = \frac{1 - \beta d}{\beta(1 - \beta)}. \quad (\text{A.2})$$

It is straightforward to prove that this score is proportional to the first Meixner polynomial since:

$$M_1(d_i; \beta) = \frac{1 - \beta d}{\sqrt{1 - \beta}}, \quad (\text{A.3})$$

and so:

$$\frac{\partial \ln f_D(d; \beta)}{\partial \beta} = \frac{M_1(d_i; \beta)}{\beta \sqrt{1 - \beta}}. \quad (\text{A.4})$$

Consequently, the orthonormal polynomials with degrees greater or equal to 2 are also proportional to the score. Bontemps (2006) shows that in such a case, the moments  $M_1(d_i; \beta)$  are robust to parameter uncertainty. Robust moments, defined by the projection of the moments on the score function, correspond exactly to the initial moments. So, it is possible to use the conditions  $M_j(d_i; \hat{\beta})$  in spite of  $M_j(d_i; \beta)$  in the definition of the GMM statistic  $J_{IND}$ , without any change in the asymptotic distribution as soon as  $\hat{\beta}$  is a square- $N$ -root-consistent estimator of the true violation rate  $\beta$ .

## B Appendix: Dufour (2006) Monte-Carlo Method

To implement this technique we first generate  $M$  independent realizations of the test statistic,  $S_i$ ,  $i = 1, \dots, M$ , under the null hypothesis, *i.e.*, using durations constructed from independent Bernoulli hit sequences. We denote by  $S_0$  the value of the test statistic obtained for the original sample. As shown by Dufour (2006), in a general case, the Monte Carlo  $p$ -values can be calculated as follows:

$$\hat{p}_M(S_0) = \frac{M \hat{G}_M(S_0) + 1}{M + 1} \quad (\text{B.1})$$

where  $\hat{G}_M(x) = 1/M \sum_{i=1}^M \mathbb{I}(S_i \geq x)$  and  $\mathbb{I}(\cdot)$  is the indicator function. When it is possible for a given simulation of the test statistic (under  $H_0$ ) to find the same value of  $S$  for at least two simulations, *i.e.*  $\Pr([S_i = S_j]) \neq 0$ , the function of empirical survival has be written as follows:

$$\tilde{G}_M(S_0) = 1 - \frac{1}{M} \sum_{i=1}^M \mathbb{I}(S_i \leq S_0) + \frac{1}{M} \sum_{i=1}^M \mathbb{I}(S_i = S_0) \times \mathbb{I}(U_i \geq U_0), \quad (\text{B.2})$$

where  $U_i$ ,  $i = 0, 1, \dots, M$  follows a uniform distribution on  $[0, 1]$ . So, in order to calculate the empirical power of test statistic  $J_{CC}(p)$ , we just need to simulate under  $H_0$ ,  $M$  independent realizations of the GMM test statistics denoted  $J_{CC,1}(p)$ ,  $J_{CC,2}(p)$ , ...,  $J_{CC,M}(p)$ . If we note  $J_{CC,0}(p)$  the test statistic obtained under  $H_1$  (by using Historical Simulation method), we reject  $H_0$  if:

$$\hat{p}_M[J_{CC,0}(p)] = \frac{M \tilde{G}_M[J_{CC,0}(p)] + 1}{M + 1} \leq a. \quad (\text{B.3})$$

where  $a$  denotes the nominal size. Statistic  $J_{CC}(p)$  simulated  $N$  times under  $H_1$  gives the test power, equal to the number of times when  $\hat{p}_M[J_{CC}(p)]$  is inferior or equal to the nominal size  $a$  (fixed at 10% in our replications), divided by  $N$  simulations. The nominal size of the test is thus respected for finite size sample. In our application, we set  $M$  at 9,999. This method allows the comparison of the empirical power of various tests independently from their empirical size.



## C References

BERKOWITZ, J. , CHRISTOFFERSEN, P. F. AND PELLETIER, D. (2005), "Evaluating Value-at-Risk models with desk-level data", *Working Paper*, University of Houston.

BONTEMPS, C. (2006), "Moment-based tests for discrete distributions", *Working Paper*.

BONTEMPS, C. AND N. MEDDAHI (2005), "Testing normality: A GMM approach", *Journal of Econometrics*, 124, pp. 149-186.

BONTEMPS, C. AND N. MEDDAHI (2006), "Testing distributional assumptions: A GMM approach", *Working Paper*.

CAMPBELL, S. D. (2007), "A review of backtesting and backtesting procedures", *Journal of Risk*, 9(2), pp 1-18.

CHRISTOFFERSEN, P. F. (1998), "Evaluating interval forecasts", *International Economic Review*, 39, pp. 841-862.

CHRISTOFFERSEN, P. F. AND D. PELLETIER (2004), "Backtesting Value-at-Risk: A duration-based approach", *Journal of Financial Econometrics*, 2, 1, pp. 84-108.

DUFOUR, J.-M. (2006), "Monte Carlo tests with nuisance parameters: a general approach to finite sample inference and nonstandard asymptotics", *Journal of Econometrics*, vol. 127(2), pp. 443-477.

ENGLE, R. F., AND MANGANELLI, S. (2004), "CAViaR: Conditional Autoregressive Value-at-Risk by regression quantiles", *Journal of Business and Economic Statistics*, 22, pp. 367-381.

ESCANCIANO JC AND OLMO J (20074), "Estimation risk effects on backtesting for parametric Value-at-Risk models", *Center for Applied Economics and Policy Research*, Working Paper, 05.

JORION, P. (2007), *Value-at-Risk*, Third edition, McGraw-Hill.

HAAS, M. (2005), "Improved duration-based backtesting of Value-at-Risk",

*Journal of Risk*, 8(2), pp. 17-36.

HANSEN L.P., (1982), "Large sample properties of Generalized Method of Moments estimators", *Econometrica*, 50, pp. 1029–1054.

KUPIEC, P.. (1995), "Techniques for verifying the accuracy of risk measurement models", *Journal of Derivatives*, 3, pp. 73-84.

NAKAGAWA, T., AND OSAKI, S. (1975), "The discrete Weibull distribution", *IEEE Transactions on Reliability*, R-24, pp 300–301.

Table 1. Empirical Size of 10% Asymptotic CC Tests

Backtesting 1% VaR					
Sample	$J_{UC}$	$J_{CC} (2)$	$J_{CC} (3)$	$J_{CC} (5)$	$LR_{CC}$
$T = 250$	0.0387	0.0237	0.0142	0.0066	0.1090
$T = 500$	0.0535	0.0253	0.0168	0.0099	0.1809
$T = 750$	0.0701	0.0486	0.0382	0.0283	0.1605
$T = 1000$	0.0867	0.0684	0.0528	0.0401	0.1598
$T = 1250$	0.0834	0.0653	0.0502	0.0401	0.1324
$T = 1500$	0.0925	0.0736	0.0589	0.0456	0.1316
Backtesting 5% VaR					
Sample	$J_{UC}$	$J_{CC} (2)$	$J_{CC} (3)$	$J_{CC} (5)$	$LR_{CC}$
$T = 250$	0.0786	0.0615	0.0489	0.0402	0.1381
$T = 500$	0.0942	0.0759	0.0558	0.0460	0.1349
$T = 750$	0.0979	0.0858	0.0691	0.0534	0.1431
$T = 1000$	0.0955	0.0814	0.0684	0.0521	0.1472
$T = 1250$	0.0944	0.0853	0.0701	0.0512	0.1556
$T = 1500$	0.1002	0.0840	0.0692	0.0515	0.1719

Notes: Under the null, the hit data are i.i.d. from a Bernoulli distribution. The results are based on 10,000 replications. For each sample, we provide the percentage of rejection at a 10% level.  $J_{cc}(p)$  denotes the GMM based conditional coverage test with  $p$  moment conditions.  $J_{uc}$  denotes the unconditional coverage test obtained for  $p=1$ .  $LR_{cc}$  denotes the Weibull conditional coverage test proposed by Berkowitz et al. (2005).

Table 2. Fraction of Samples where Tests are Feasible

Size Simulations				
Sample Size	1% VaR		5% VaR	
	$J_{CC}$	$LR_{CC}$	$J_{CC}$	$LR_{CC}$
$T = 250$	0.9217	0.6249	1.000	0.9999
$T = 500$	0.9944	0.9326	1.000	1.000
$T = 750$	0.9996	0.9917	1.000	1.000
$T = 1000$	1.0000	0.9984	1.000	1.000
Power Simulations ( $Te = 500$ )				
Sample Size	1% VaR		5% VaR	
	$J_{CC}$	$LR_{CC}$	$J_{CC}$	$LR_{CC}$
$T = 250$	0.7953	0.5972	0.9905	0.9681
$T = 500$	0.9625	0.8894	0.9999	0.9995
$T = 750$	0.9970	0.9852	1.000	1.000
$T = 1000$	0.9691	0.9985	1.000	1.000

Notes: The results are based on 10,000 replications. For each sample and for each test, we provide the percentage of samples for which the statistic can be computed.  $J_{cc}$  denotes the GMM based (un)conditional coverage test. For the J test, note that the feasible ratios are independent of the number  $p$  of moments used.  $LR_{cc}$  denotes the Weibull conditional coverage test proposed by Berkowitz et al. (2005).

Table 3. Power of 10% Finite Sample Tests on 1% VaR

Length of Rolling Estimation Window $Te = 250$					
Sample	$J_{UC}$	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	$LR_{CC}$
$T = 250$	0.4132	0.4369	0.4580	0.4980	0.2098
$T = 500$	0.3990	0.4674	0.5140	0.5710	0.3244
$T = 750$	0.3924	0.4897	0.5626	0.6276	0.4268
$T = 1000$	0.4416	0.5496	0.6396	0.7115	0.5529
$T = 1250$	0.4924	0.6062	0.6989	0.7752	0.6728
$T = 1500$	0.5339	0.6592	0.7563	0.8284	0.7465
Length of Rolling Estimation Window $Te = 500$					
Sample	$J_{UC}$	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	$LR_{CC}$
$T = 250$	0.4329	0.4554	0.4790	0.5177	0.1913
$T = 500$	0.4074	0.4837	0.5359	0.5858	0.3570
$T = 750$	0.3749	0.5320	0.5967	0.6623	0.4917
$T = 1000$	0.3567	0.5767	0.6543	0.7170	0.6014
$T = 1250$	0.3532	0.6389	0.7387	0.7935	0.7121
$T = 1500$	0.3679	0.6876	0.7900	0.8410	0.7772

Notes: The results are based on 10,000 replications. For each sample, we provide the percentage of rejection at a 10% level.  $J_{CC}(p)$  denotes the GMM based conditional coverage test with  $p$  moment conditions.  $J_{UC}$  denotes the unconditional coverage test obtained for  $p=1$ .  $LR_{CC}$  denotes the Weibull conditional coverage test proposed by Berkowitz et al. (2005).

Table 4. Power of 10% Finite Sample Tests on 5% VaR

Length of Rolling Estimation Window $Te = 250$					
Sample	$J_{UC}$	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	$LR_{CC}$
$T = 250$	0.3956	0.5738	0.6106	0.6100	0.3652
$T = 500$	0.2791	0.7266	0.7865	0.7792	0.4920
$T = 750$	0.2656	0.8432	0.8935	0.8875	0.6274
$T = 1000$	0.2462	0.9092	0.9467	0.9467	0.7439
$T = 1250$	0.2574	0.9513	0.9737	0.9719	0.8251
$T = 1500$	0.2534	0.9731	0.9873	0.9848	0.8700
Length of Rolling Estimation Window $Te = 500$					
Sample	$J_{UC}$	$J_{CC}(2)$	$J_{CC}(3)$	$J_{CC}(5)$	$LR_{CC}$
$T = 250$	0.4282	0.5880	0.6325	0.6260	0.4284
$T = 500$	0.4224	0.7857	0.8303	0.8310	0.6013
$T = 750$	0.3512	0.8738	0.9131	0.9076	0.7014
$T = 1000$	0.2953	0.9312	0.9589	0.9572	0.7924
$T = 1250$	0.2663	0.9618	0.9799	0.9785	0.8667
$T = 1500$	0.2397	0.9802	0.9920	0.9900	0.9071

Notes: The results are based on 10,000 replications. For each sample, we provide the percentage of rejection at a 10% level.  $J_{CC}(p)$  denotes the GMM based conditional coverage test with  $p$  moment conditions.  $J_{UC}$  denotes the unconditional coverage test obtained for  $p=1$ .  $LR_{CC}$  denotes the Weibull conditional coverage test proposed by Berkowitz et al. (2005).

Table 5. Backtesting Tests of 5% VaR Forecasts for Nasdaq Index

Backtesting Tests	VaR Forecasting Methods			
	Statistic	GARCH-t( $d$ )	HS	CAViaR
<i>Unconditional Coverage</i>	Hits Freq.	0.036	0.036	0.028
	$J_{UC}$	1.197 (0.261)	1.197 (0.261)	4.066 (0.036)
<i>Independence Tests</i>	$J_{IND}(2)$	0.248 (0.646)	0.186 (0.719)	0.080 (0.867)
	$J_{IND}(4)$	0.383 (0.779)	4.652 (0.036)	0.532 (0.663)
	$J_{IND}(6)$	0.385 (0.907)	7.857 (0.016)	2.127 (0.299)
	$LR_{IND}$	0.512 (0.532)	1.772 (0.236)	0.058 (0.823)
<i>Conditional Coverage</i>	$J_{CC}(2)$	1.522 (0.315)	2.708 (0.157)	4.466 (0.071)
	$J_{CC}(4)$	1.545 (0.522)	11.14 (0.025)	9.874 (0.030)
	$J_{CC}(6)$	1.586 (0.646)	11.89 (0.030)	12.05 (0.031)
	$LR_{CC}$	2.335 (0.364)	3.596 (0.213)	4.198 (0.168)

Notes: the hit empirical frequency is the ratio of VaR violations to the sample size ( $T=250$ ) observed for the Nasdaq between June 22, 2005 and June 20, 2006. Three methods of VaR forecasting are used: the GARCH with Student conditional, distribution, historical simulation (HS) and a CAViaR (Engle and Manganelli, 2004). For VaR method,  $J_{uc}$  denotes the unconditional coverage test statistic obtained for  $p=1$ .  $J_{ind}(p)$  and  $J_{cc}(p)$  respectively denote the GMM based independence and conditional coverage tests base on  $p$  moments conditions. The number of moments is fixed to 2, 4 or 6.  $LR_{ind}$  and  $LR_{cc}$  respectively denote the Weibull independence and conditional coverage tests respectively proposed by Christoffersen and Pelletier (2004) and Berkowitz et al. (2005). For all these tests, the numbers in parenthesis denote the corresponding  $p$ -values corrected by the Dufour's Monte Carlo procedure (Dufour, 2006).

Fig. C.1. GARCH- $t(v)$  Simulated Returns with 1% and 5% VaR from HS ( $Te = 250$ ).

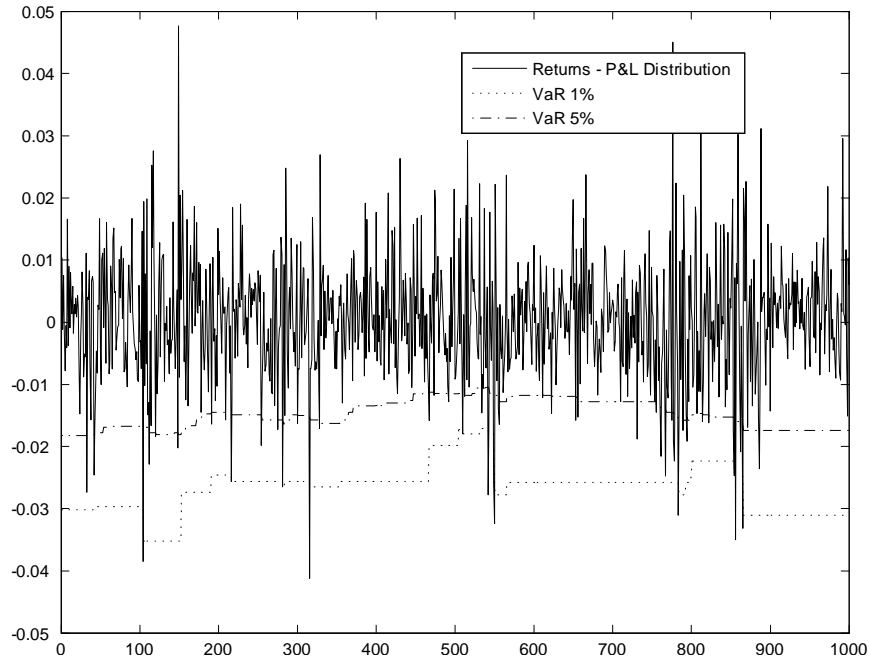


Fig. C.2. Empirical Power: Sensitivity Analysis to the choice of  $p$

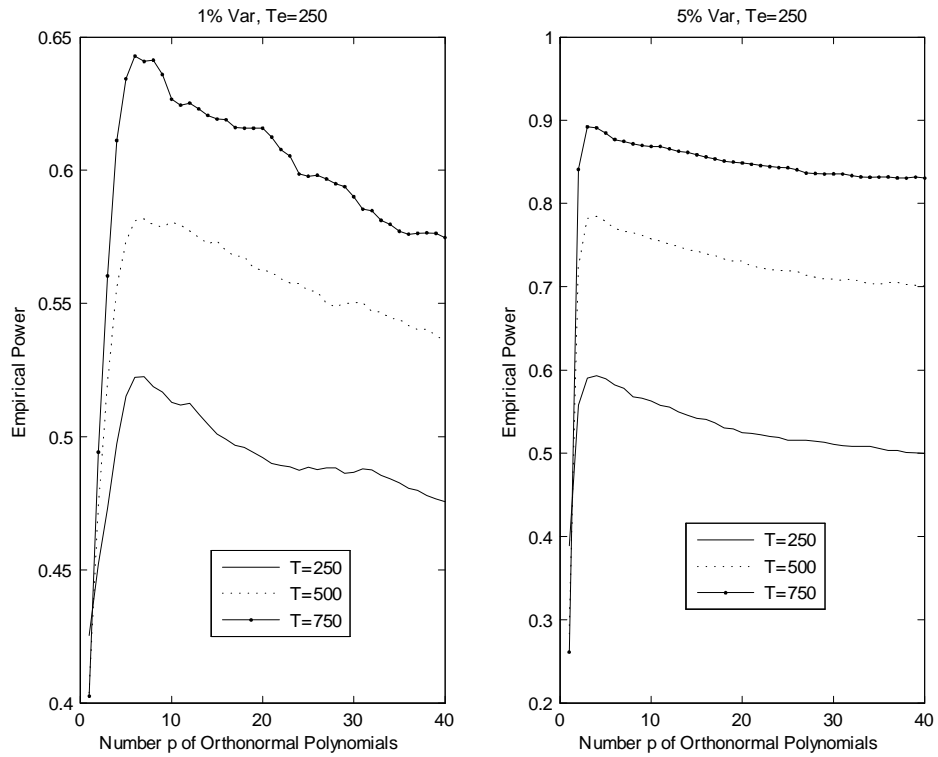




Fig. C.3. Historical Returns and 5% VaR Forecasts. Nasdaq (June 2005- June 2006)

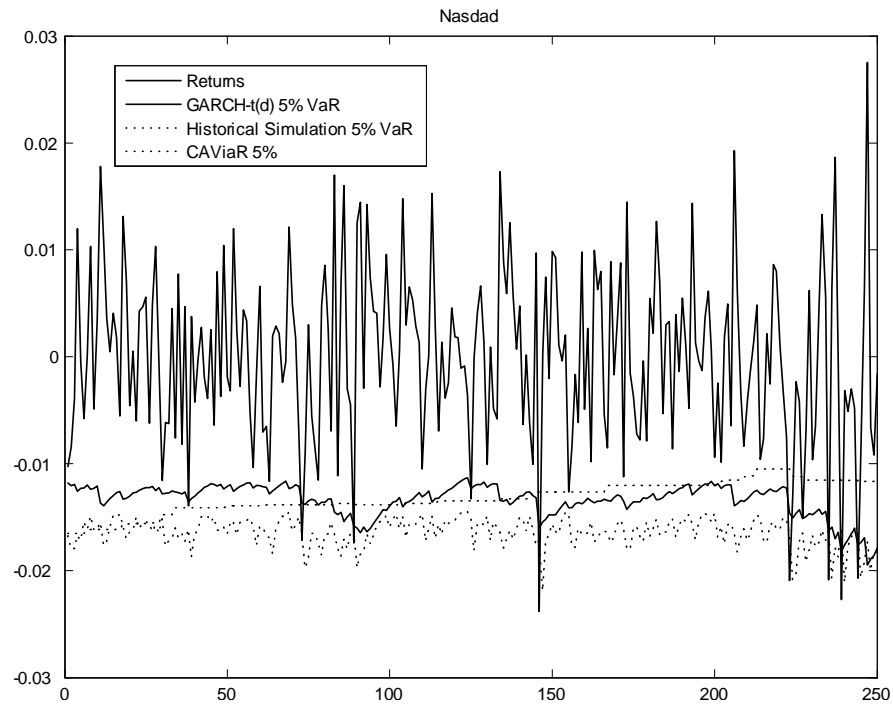


Fig. C.4. 5%VaR Violations. Nasdaq (June 2005 - June 2006,  $T = 250$ )

